Crime Analysis

Final Project

Yiru Fei

12/1/2019

Introduction

Stories about crime have always been popular in the news media as well as film makers, and in the modern time, thousands of crimes are committed every day. A city where has a higher population may have more crimes occur. San Fansico is the fourth-most populous in California, with 883,305 residents as of 2018. And in 2011, 50 murders were reported, which is 6.1 per 100,000 people. There were about 4,210 motor vehicle thefts, 3,142 robberies, and 2,139 assaults. Boston, with an estimated population of 694,583 in 2018, is the most populous city in New England.

For this project, I will compare the crime reports between Boston and San Fansico, and try to find the differences and similarities of crime categories between these two cities by using EDA, text analysis, and mapping.

Data Description

The dataset for Boston is the crime incident reports, which are provided by Boston Police Department. This is a dataset containing records from the new crime incident report system, which includes a reduced set of fields focused on capturing the type of incident as well as when and where it occurred. Records in the new system begin in June of 2015. The dataset for San Francisco includes police incident reports filed by officers and by individuals through self-service online reporting for non-emergency cases. Reports included are those for incidents that occurred starting January 1, 2018 onward and have been approved by a supervising officer.

Due to the time span of the two data is different, I only use the crime reports from 2018 to 2019. After removing those missing data, I have 159676 observations for reports of Boston, and 213438 observations for reports of San Francisco. The following are the variables and descriptions used in this project.

-Incident Datetime: The date and time when the incident occurred

-Incident Number: The number issued on the report

-Incident Code: System codes to describe a type of incident

-Incident Category: A category mapped on to the Incident Code used in statistics reporting

-Street: Street names that closest to the original incident

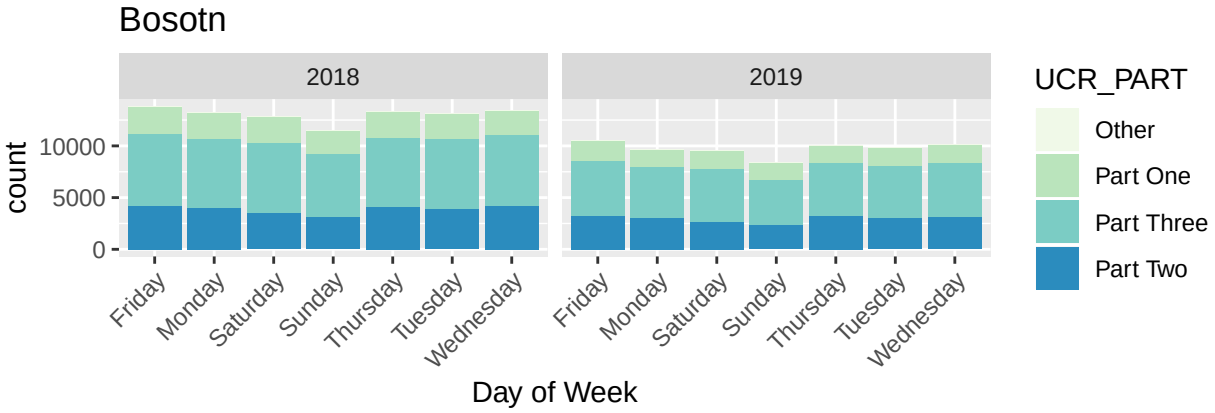
-Type Description: The description of the report type

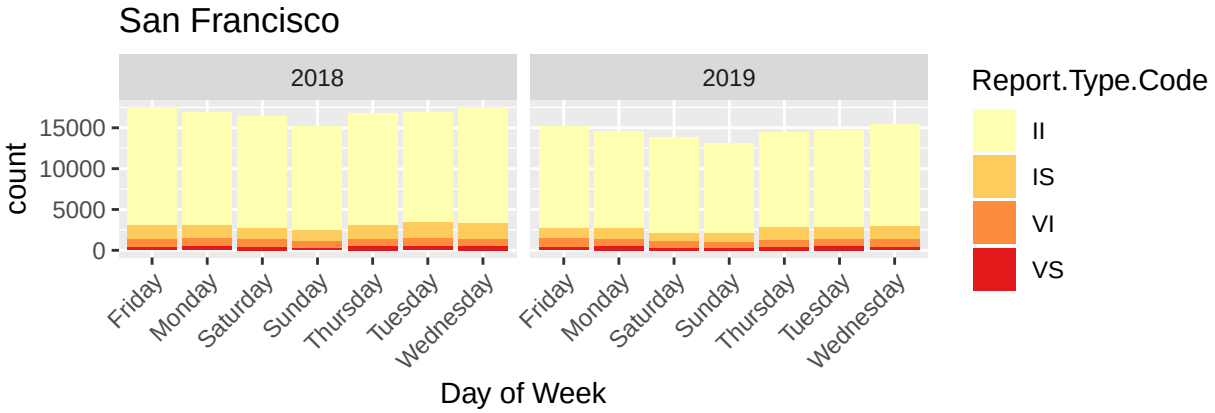
-Latitude: Latitude of incident

-Longitude: Longitude of incident

EDA

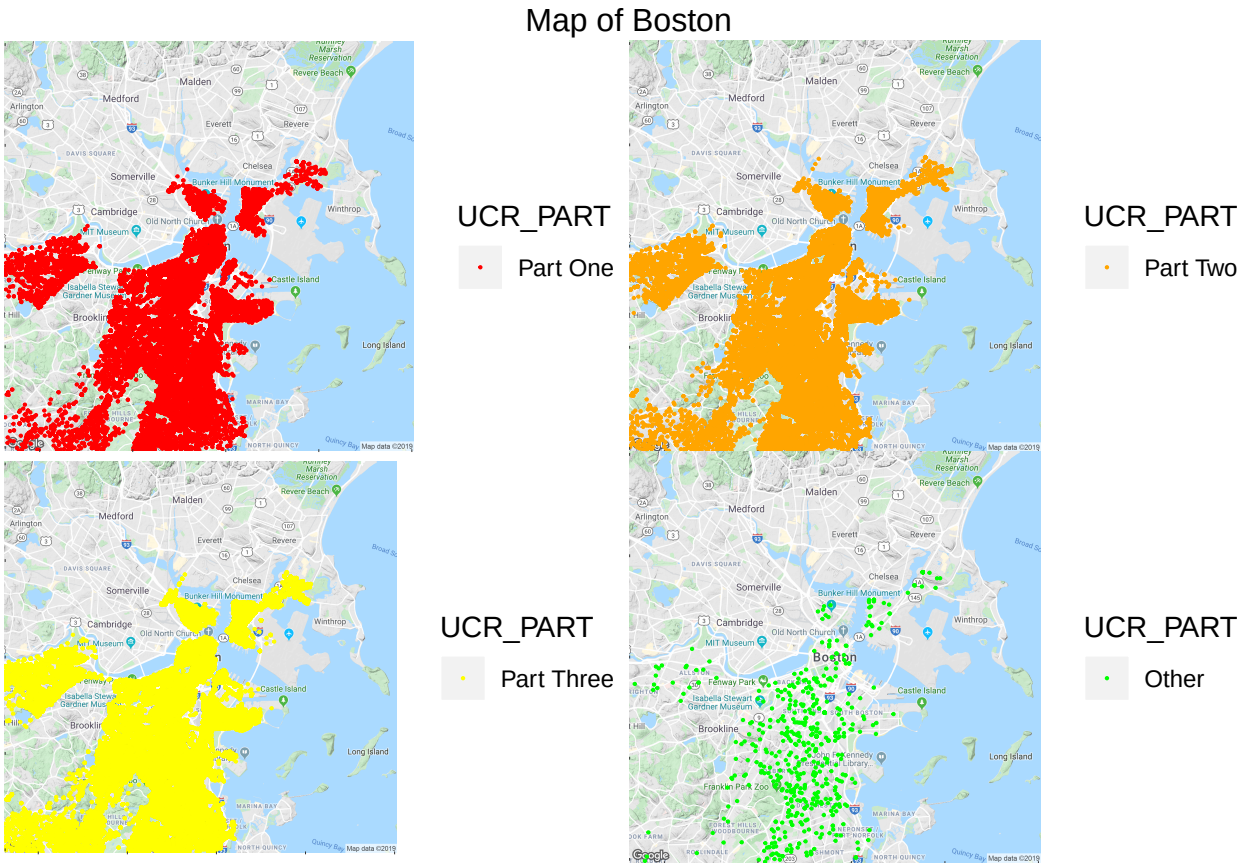
1.Bar Plot





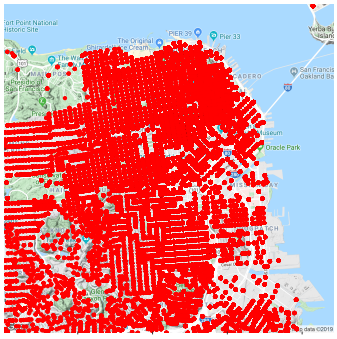
These four bar-plots show the relationships between the day of the week, year, and crime types. Fist of all, I find that Sunday has the least crime report in both Boston and San Francisco. In both cities, crimes are less in 2019 than in 2018. Furthermore, in Boston, the most reported crimes are type three crime, but the most reported crimes are II, which is the initial crime in San Francisco. Finally, from the total number of crimes on a different day of the week, I find that San Francisco has more crime than Boston since the highest number of the y-axis of San Francisco is over 15000. Compared to the amount of reported crime, Boston has a much lower criminal record in 2019 than in 2018, this phenomenon maybe because of the incomplete data of 2019.

2.Mapping



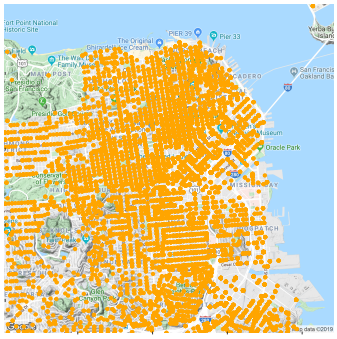
The map of different crime types for Boston shows that most crimes occur around Boston Common and South End. The farther away from the city center, the fewer crimes occur, so Jamaica Plain is a relatively safer place.

Map of San Franciscon



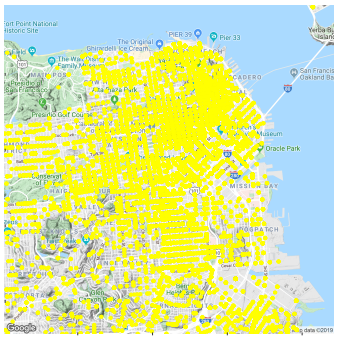
Report.Type.Code

II



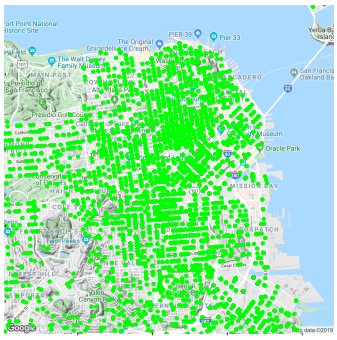
Report.Type.Code

VI



Report.Type.Code

IS

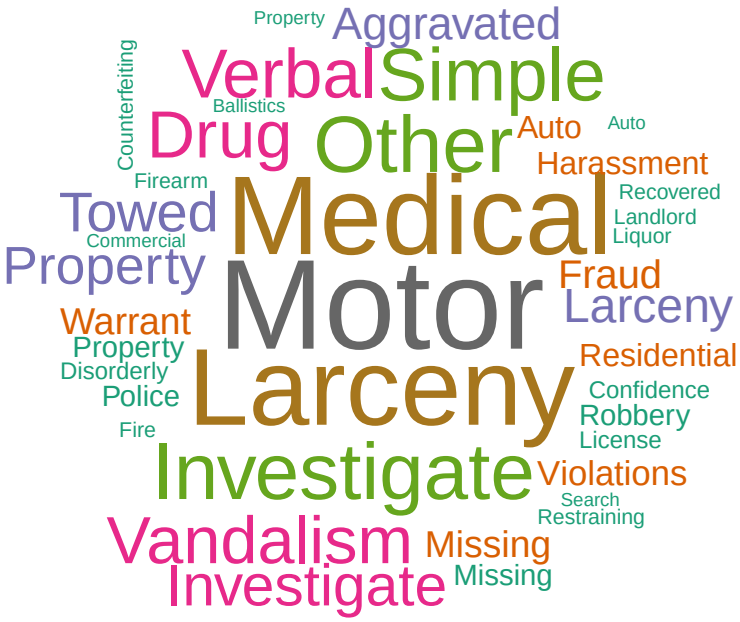


Report.Type.Code

VS

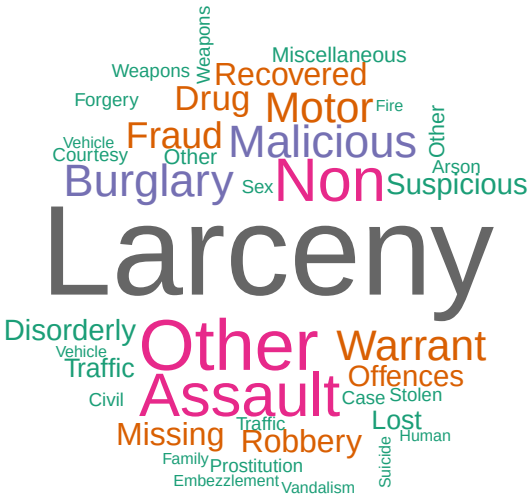
The map of different crime types for San Francisco shows that most crimes occur around Tenderloin and North Beach. Unlike Boston, the crimes that happened in San Francisco were distributed evenly. The only place where has fewer points than others is a park, which means the population density is lower than in other areas.

3.Text Analysis



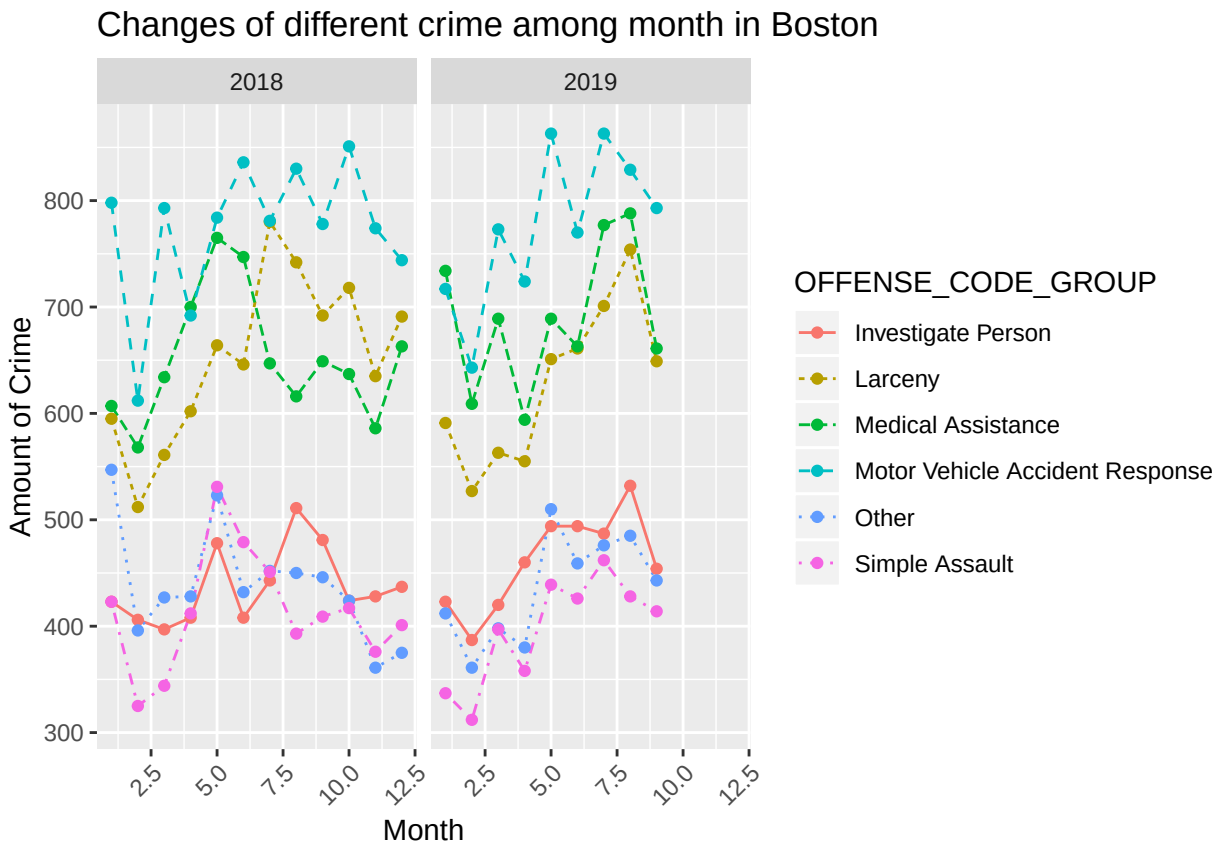
The word cloud of crimes in Boston indicates that motor vehicle accident response, medical assistance, and

larceny are the three most common crimes in Boston.



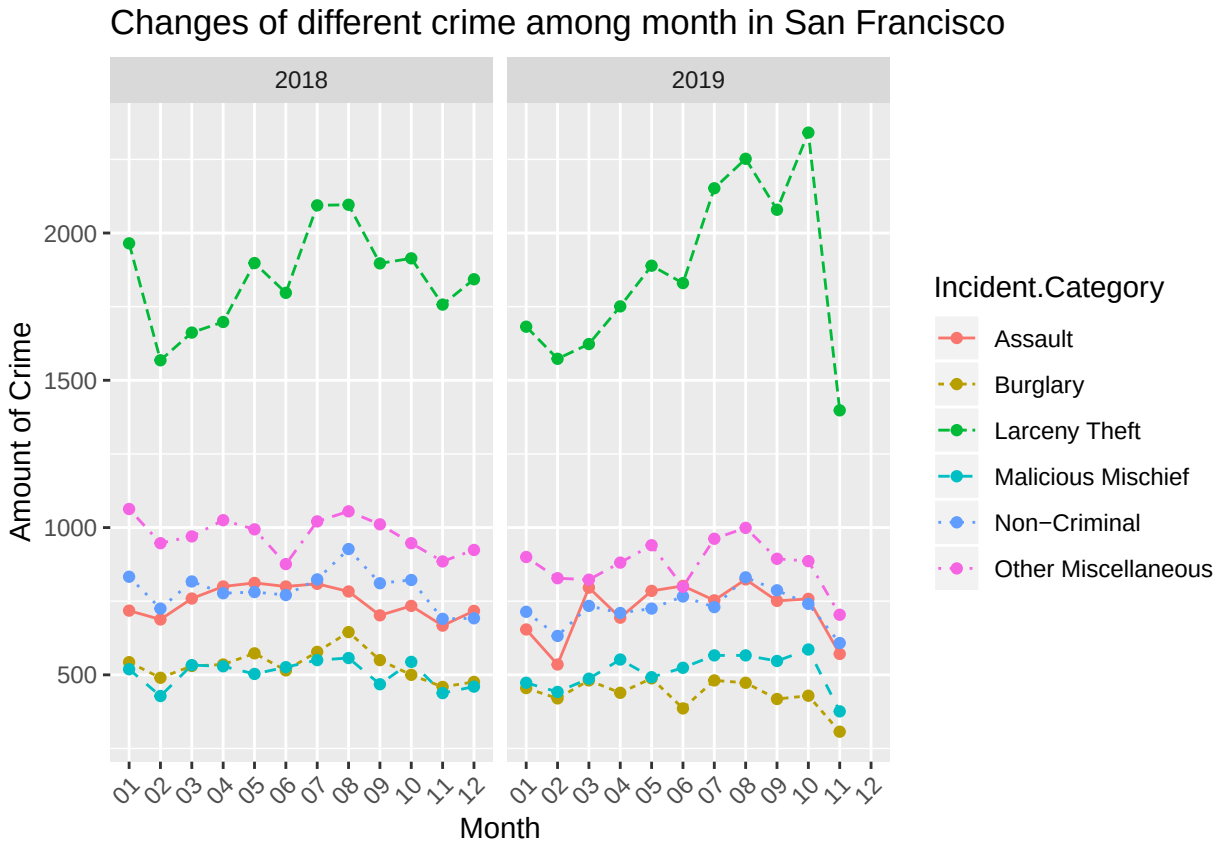
The word cloud of crimes in San Francisco shows that larceny theft and other miscellaneous are the most common crimes in San Francisco. Furthermore, larceny theft occurs twice as often as other miscellaneous. The data shows that the total records of larceny theft are 41948, and the total records of other miscellaneous are 20835.

4.Line Plor



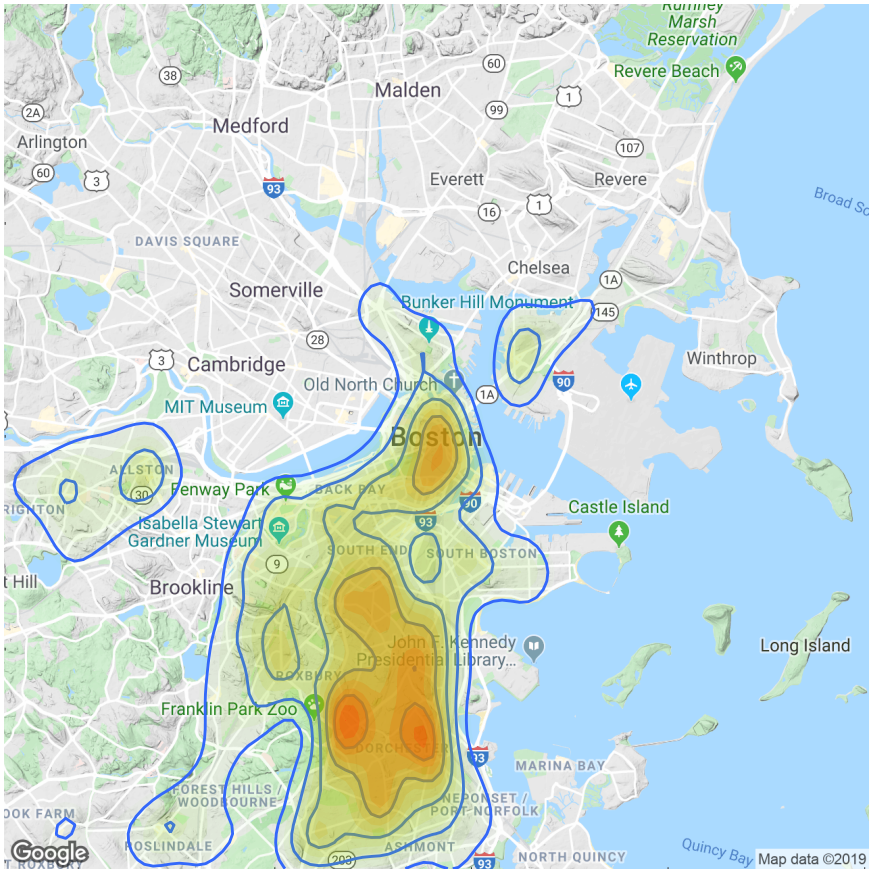
This line plot shows the changes in the most six common crime in Boston. From the plot, I find reported crimes are hight in May both in 2018 and in 2019, and reported crimes are low in February. The frequency of these six crimes has not changed much between these two years. Therefore, based on the first barplot, the

number of criminal records decreased may be caused by the decreasing of other rare crime types.

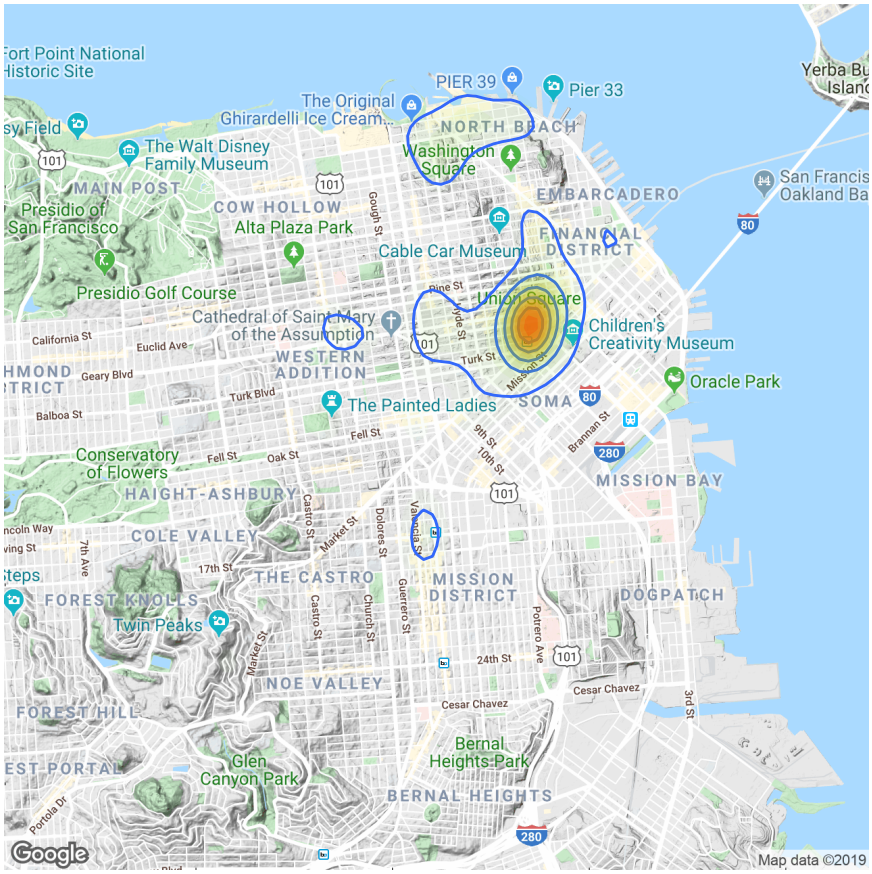


Same as Boston, the reported crimes of San Francisco are low in February, but the relatively high records of crimes is in August. Furthermore, the sudden drop in crime rates in November 2019 should also be due to incomplete data. From the line plot of San Francisco, I can get the same information as from the word cloud, larceny theft occurs twice as often as other miscellaneous.

5.Mapping



This map displays the frequency of motor vehicle accident response in Boston. The darker the color means the higher the rate of occurrence. Similar to the previous maps, Boston Common and South End have more reported crimes than other areas.



This map shows the frequency of most common crime, larceny-theft, in San Francisco. Unexpectedly, almost all the larceny-theft crimes happen around Tenderloin, and a small percentage of crimes occur around North Beach.

Conclusion

-Similarities:
Both cities have the lowest crime rates on Sunday and relatively few criminal records in February than the majority of other months. Furthermore, crime rates will be higher in densely populated areas, and the crime rate is lower near the park, such as Fenway Park and Forest Knolls.

-Differences:
Based on the record of the type of crimes, three main types of the crime of Boston are closed. However, for San Francisco, the number of initial crimes is much more than the other three types. Most crimes in Boston occur in Downtown and South End, and the frequency of crimes in South End is higher and broader than in Downtown. But, most crimes happened close to downtown in San Francisco, and crime rates are low in other areas.

References

https://en.wikipedia.org/wiki/San_Francisco
https://en.wikipedia.org/wiki/Boston
https://data.boston.gov/dataset/crime-incident-reports-august-2015-to-date-source-new-system
https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-2018-to-Present/wg3w-h783